A BLIND MONTE CARLO DETECTION-ESTIMATION METHOD FOR OPTICAL COHERENCE TOMOGRAPHY

Georg Kail[a], Clemens Novak[a], Bernd Hofer[b], and Franz Hlawatsch[a]

[a]Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology
Gusshausstrasse 25/389, A-1040 Vienna, Austria; e-mail: gkail@nt.tuwien.ac.at
[b]School of Optometry & Vision Sciences, Cardiff University
Mandy Road, Cardiff CF24 4LU, Wales, UK; e-mail: hoferb@cf.ac.uk

ABSTRACT

We consider the parametric analysis of frequency-domain optical coherence tomography (OCT) signals. A Monte Carlo (Gibbs sampler) detection-estimation method for determining the depths and reflection coefficients of tissue interfaces (reflective sites in the tissue) is proposed. Our method is blind since it estimates the instrumentation-dependent "fringe" function along with the tissue parameters. Sparsity of the detected interfaces is enforced by an impulse detector and a modified Bernoulli-Gaussian prior with a minimum distance constraint. Numerical results using synthetic and real signals demonstrate the excellent performance and fast convergence of our method.

Index Terms— Optical coherence tomography, Bayesian analysis, Gibbs sampler, Monte Carlo method, detection, estimation, Bernoulli-Gaussian model, blind deconvolution.

1. INTRODUCTION

In optical coherence tomography (OCT), the depth-dependent reflection properties of tissue are measured via an optical interferometer [1]. In the case of the retina (or other layered tissues), the OCT signal transformed into the time domain can be approximated by the convolution of a sparse impulse train (the "retina function," abbreviated RF) with a pulse (the "fringe"). Each RF impulse corresponds to a tissue interface (reflective site in the tissue), while the fringe depends on the instrumentation and determines the imaging resolution.

Here, we propose a Monte Carlo (Gibbs sampler) detection-estimation method for determining the locations and amplitudes of the RF impulses. These correspond to the depths and reflection coefficients, respectively, of the tissue interfaces and are of clinical relevance. Our method is blind in that the fringe need not be known but is estimated along with the RF parameters. It is inspired by the Gibbs sampler based blind deconvolution method proposed for seismic signals in [2], but improves on that method due to its much faster convergence and lack of spurious impulses. These improvements are achieved by the use of a dedicated impulse detector, a sparsity-enforcing modified Bernoulli-Gaussian prior incorporating a minimum distance constraint, and a parametric model for the fringe.

Two alternative methods that have been proposed for OCT are the SMLR detection method [3,4] and a cepstral deconvolution method [5]. The SMLR method requires the fringe to be known and may converge to a local maximum of the posterior (this does not happen with Gibbs sampler methods). The cepstral method does not perform impulse detection; furthermore, it does not consider additive measurement noise and does not compensate for modeling errors.

This paper is organized as follows. The signal model and prior distributions are discussed in Sections 2 and 3, respectively. Section 4 describes the Monte Carlo detection-estimation method. The distributions used by the Gibbs sampler are developed in Section 5. Finally, numerical results are presented in Section 6.

2. SIGNAL MODEL

Convolution model. The sampled signal measured by a frequency-domain OCT device [1] can be modeled as (cf. [5])

$$\tilde{Y}_l = |1 + R_l|^2 F_l + N_l = \left(1 + R_l + R_l^* + |R_l|^2\right)F_l + N_l \,,$$

where R_l and F_l are the Fourier transforms of, respectively, the RF and the fringe and N_l is measurement noise (not included in [5]). The component 1 (caused by the interferometer's reference branch) is suppressed, along with systematic modeling errors, by a standard preprocessing procedure in which the average of the signals of many local depth scans is subtracted from $\tilde{Y}_l$. Neglecting $|R_l|^2$ ($\ll |R_l|$) and performing an inverse discrete Fourier transform then yields the time-domain signal $y_k = (r_k + r_{-k}^*) * f_k + n_k$, where $*$ denotes convolution. The component $r_k * f_k$ is effectively causal and does not overlap with the anticausal component $r_{-k}^* * f_k$. Retaining a causal block of K samples, we obtain the convolution model

$$x_k = r_k * f_k + n_k \,, \quad k = 0, \dots, K-1 \,.$$

Retina function. The RF r_k consists of irregularly spaced impulses with complex amplitudes. This can be modeled as

$$r_k = a_k b_k \,, \quad k = 0, \dots, K-1 \,,$$

where the "impulse indicator" $b_k \in \{0, 1\}$ indicates the impulse positions and $a_k \in \mathbb{C}$ describes the impulse amplitudes for those k where $b_k = 1$ (the a_k are irrelevant elsewhere). Our goal is to detect the impulse positions (i.e., the k for which $b_k = 1$) and to estimate the corresponding amplitudes a_k, for an *unknown* fringe f_k.

Denoting by $\mathbf{x}, \mathbf{r}, \mathbf{b}, \mathbf{a}, \mathbf{f}$, and $\mathbf{n}$ the length-K vectors corresponding to the sequences x_k, r_k, b_k, a_k, f_k, and n_k, respectively, the signal $x_k = (a_k b_k) * f_k + n_k$ can be written as

$$\mathbf{x} = \mathbf{Fr} + \mathbf{n} = \mathbf{FBa} + \mathbf{n} \,, \tag{1}$$

where $\mathbf{F} \triangleq \text{toep}(\mathbf{f})$ is the $K \times K$ Toeplitz matrix generated by $\mathbf{f}$ and $\mathbf{B} \triangleq \text{diag}(\mathbf{b})$ is the $K \times K$ diagonal matrix with main diagonal $\mathbf{b}$.

Fringe. We represent the fringe $\mathbf{f}$ by an expansion

$$\mathbf{f} = \sum_{l=1}^{L} \alpha_l \mathbf{h}_l = \mathbf{H}\boldsymbol{\alpha} \,, \tag{2}$$

where the length-K vectors $\mathbf{h}_l$, $l = 1, \dots, L$ correspond to the first L ($\ll K$) Hermite functions, $\mathbf{H} \triangleq (\mathbf{h}_1 \cdots \mathbf{h}_L)$, and $\boldsymbol{\alpha} \triangleq (\alpha_1 \cdots \alpha_L)^T$. The Hermite functions cover an elliptic time-frequency region [6,7]. Their number L and time-frequency scaling are chosen depending on the effective duration and bandwidth of the fringe.

Defining the $K \times K$ Toeplitz matrix $\mathbf{R} \triangleq \text{toep}(\mathbf{r})$, we have $\mathbf{Fr} = \text{toep}(\mathbf{r})\,\mathbf{f} = \mathbf{Rf}$, so that (1) can be written as

$$\mathbf{x} = \mathbf{Rf} + \mathbf{n} = \mathbf{RH}\boldsymbol{\alpha} + \mathbf{n} \,. \tag{3}$$

3. STOCHASTIC MODEL

The Bayesian methodology requires prior distributions for all quantities to be estimated or detected. In particular, we use a modified Bernoulli-Gaussian prior for the RF $r_k = a_k b_k$, as described next.

Impulse indicator. To develop the prior of the impulse indicator vector $\mathbf{b}$, we first consider all $b_k \in \{0, 1\}$ to be independent and identically distributed (iid) Bernoulli variables. This corresponds to the auxiliary probability mass function (pmf) $\tilde{p}(\mathbf{b}) = \prod_{k=0}^{K-1} \tilde{p}(b_k)$, with $\tilde{p}(b) = p_0$ for $b = 0$ and $\tilde{p}(b) = 1 - p_0$ for $b = 1$. However, to exclude closely spaced impulses (which are not physically relevant and would lead to spurious impulses in the detection result), we enforce a minimum distance $d_{\min}$ between any two nonzero entries $b_k = 1$ and $b_l = 1$. Let $\mathcal{C}$ be the set of all $\mathbf{b}$ satisfying this *minimum distance condition*, and let $\mathrm{I}_\mathcal{C}(\mathbf{b})$ be 1 if $\mathbf{b} \in \mathcal{C}$ and 0 otherwise. Then, the prior pmf of $\mathbf{b}$ is chosen (up to an irrelevant normalization) as

$$p(\mathbf{b}) \propto \mathrm{I}_\mathcal{C}(\mathbf{b})\, \tilde{p}(\mathbf{b}) = \begin{cases} \tilde{p}(\mathbf{b}), & \mathbf{b} \in \mathcal{C} \\ 0, & \text{otherwise.} \end{cases}$$

Note that the b_k are no longer statistically independent.

Amplitudes. The amplitudes a_k are modeled as iid zero-mean circularly complex Gaussian, i.e., the prior probability density function (pdf) of $\mathbf{a}$ is $p(\mathbf{a}) = \mathcal{CN}(\mathbf{0}, \sigma_a^2 \mathbf{I})$, with σ_a^2 suitably chosen. Thus, $|a_k|$ is Rayleigh-distributed, which means that amplitudes around zero have a small probability. This helps avoid small impulses (which are insignificant and typically artifacts) in the detection result.

Fringe coefficients. The fringe coefficient vector $\boldsymbol{\alpha}$ in (2) is modeled as $p(\boldsymbol{\alpha}) = \mathcal{CN}(\mathbf{0}, \sigma_\alpha^2 \mathbf{I})$.

Noise. The noise is modeled as $p(\mathbf{n}) = \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I})$. Its variance σ_n^2 is treated as a random hyperparameter and estimated. The prior of σ_n^2 is chosen as an inverse gamma distribution [8]:

$$p(\sigma_n^2) = \mathcal{IG}(\xi, \eta) = \frac{\eta^\xi}{\Gamma(\xi)} (\sigma_n^2)^{-\xi-1} e^{-\eta/\sigma_n^2} \mathrm{u}(\sigma_n^2) \quad (4)$$

($\mathrm{u}(x)$ is the unit step function). This is a conjugate prior of the Gaussian likelihood function [8], which will be useful in Section 5.

Joint prior and posterior. The quantities $\mathbf{b}$, $\mathbf{a}$, $\boldsymbol{\alpha}$, and σ_n^2 to be detected or estimated are modeled as independent. Thus, their joint prior is $p(\mathbf{b}, \mathbf{a}, \boldsymbol{\alpha}, \sigma_n^2) = p(\mathbf{b})\, p(\mathbf{a})\, p(\boldsymbol{\alpha})\, p(\sigma_n^2)$, with $p(\mathbf{b})$ etc. as discussed above. Using (1), the likelihood function $p(\mathbf{x}|\mathbf{b}, \mathbf{a}, \boldsymbol{\alpha}, \sigma_n^2)$ is obtained by setting $\mathbf{n} = \mathbf{x} - \mathbf{F}\mathbf{B}\mathbf{a}$ in $p(\mathbf{n}) = \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I})$:

$$p(\mathbf{x}|\mathbf{b}, \mathbf{a}, \boldsymbol{\alpha}, \sigma_n^2) = \frac{1}{(\pi \sigma_n^2)^K} \exp\left(-\frac{\|\mathbf{x} - \mathbf{F}\mathbf{B}\mathbf{a}\|^2}{\sigma_n^2}\right), \quad (5)$$

with $\mathbf{F} = \mathrm{toep}(\mathbf{H}\boldsymbol{\alpha})$ and $\mathbf{B} = \mathrm{diag}(\mathbf{b})$. For the joint posterior of $\mathbf{b}$, $\mathbf{a}$, $\boldsymbol{\alpha}$, and σ_n^2, we then obtain $p(\mathbf{b}, \mathbf{a}, \boldsymbol{\alpha}, \sigma_n^2|\mathbf{x}) \propto p(\mathbf{x}|\mathbf{b}, \mathbf{a}, \boldsymbol{\alpha}, \sigma_n^2) \times p(\mathbf{b}, \mathbf{a}, \boldsymbol{\alpha}, \sigma_n^2)$.

4. MONTE CARLO DETECTION-ESTIMATION METHOD

4.1. Detection of Impulses

Two optimum methods for detecting the impulse sequence $\mathbf{b} = (b_k)$ are the maximum a posteriori (MAP) *sequence* or *vector* detector,

$$\hat{\mathbf{b}}_{\mathrm{MAP}}(\mathbf{x}) \triangleq \arg \max_{\mathbf{b} \in \{0,1\}^K} p(\mathbf{b}|\mathbf{x}) = \arg \max_{\mathbf{b} \in \mathcal{C}} p(\mathbf{b}|\mathbf{x}), \quad (6)$$

and the MAP *component* detector, also known as "maximum posterior marginal/mode (MPM) detector" (e.g., [9]),

$$\hat{b}_{k,\mathrm{MAP}}(\mathbf{x}) \triangleq \arg \max_{b_k \in \{0,1\}} p(b_k|\mathbf{x}), \quad k = 0, \dots, K-1. \quad (7)$$

The first minimizes the sequence error probability $\mathrm{P}\{\hat{\mathbf{b}} \neq \mathbf{b}\}$, while the second minimizes the component error probability $\mathrm{P}\{\hat{b}_k \neq b_k\}$.

Monte Carlo detectors. Both $p(\mathbf{b}|\mathbf{x})$ and $p(b_k|\mathbf{x})$ can be derived from the posterior $p(\mathbf{b}, \mathbf{a}, \boldsymbol{\alpha}, \sigma_n^2|\mathbf{x})$ by marginalization. This being computationally prohibitive, we use a Monte Carlo approach, i.e., we generate a sample $\mathcal{S} \triangleq \{(\mathbf{b}, \mathbf{a}, \boldsymbol{\alpha}, \sigma_n^2)^{(m)}\}_{m=1,\dots,M}$ of realizations $(\mathbf{b}, \mathbf{a}, \boldsymbol{\alpha}, \sigma_n^2)^{(m)}$ from $p(\mathbf{b}, \mathbf{a}, \boldsymbol{\alpha}, \sigma_n^2|\mathbf{x})$ (as discussed in Section 5) and calculate a detector from this sample.

Using the sample $\mathcal{S}$, marginalizations are easily done by ignoring the undesired components of each realization $(\mathbf{b}, \mathbf{a}, \boldsymbol{\alpha}, \sigma_n^2)^{(m)}$. In particular, let $q(\mathbf{b})$ denote the relative multiplicity of $\mathbf{b} \in \{0, 1\}^K$ in $\mathcal{S}$ (i.e., the number of occurrences of $\mathbf{b}$ in $\mathcal{S}$, normalized by $|\mathcal{S}| = M$). Then $q(\mathbf{b})$ converges to $p(\mathbf{b}|\mathbf{x})$ underlying $\hat{\mathbf{b}}_{\mathrm{MAP}}(\mathbf{x})$ in (6) as M increases. Therefore, the *sample-based sequence detector*

$$\hat{\mathbf{b}}_\mathcal{S}(\mathbf{x}) \triangleq \arg \max_{\mathbf{b} \in \{0,1\}^K} q(\mathbf{b}) = \arg \max_{\mathbf{b} \in \mathcal{B}} q(\mathbf{b}),$$

where $\mathcal{B}$ is the set of $\mathbf{b} \in \{0, 1\}^K$ contained in $\mathcal{S}$ (i.e., for which $q(\mathbf{b}) \neq 0$), approximates $\hat{\mathbf{b}}_{\mathrm{MAP}}(\mathbf{x})$ for M sufficiently large. (Note that $\hat{\mathbf{b}}_\mathcal{S}(\mathbf{x})$ is simply the $\mathbf{b}$ occurring most often in $\mathcal{S}$.) However, the number $|\mathcal{B}| \leq |\mathcal{S}| = M$ of different $\mathbf{b}$ contained in $\mathcal{S}$ is usually much smaller than the number $|\mathcal{C}|$ of hypotheses $\mathbf{b} \in \mathcal{C}$ among which $\hat{\mathbf{b}}_{\mathrm{MAP}}(\mathbf{x})$ selects the best. Hence, $q(\mathbf{b})$ will be quite different from $p(\mathbf{b}|\mathbf{x})$, and thus $\hat{\mathbf{b}}_\mathcal{S}(\mathbf{x})$ may be different from $\hat{\mathbf{b}}_{\mathrm{MAP}}(\mathbf{x})$.

This problem is avoided by the sample-based version of the MAP *component* detector (7), which is given by

$$\hat{b}_{k,\mathcal{S}}(\mathbf{x}) \triangleq \arg \max_{b_k \in \{0,1\}} q(b_k), \quad k = 0, \dots, K-1.$$

Here, $q(b_k)$ is the relative multiplicity of b_k, i.e., the number of $\mathbf{b}$'s in $\mathcal{S}$ that have the given $b_k \in \{0, 1\}$ at position k, normalized by M. The limited sample size is no problem here since there are only two possible hypotheses for b_k. However, $\hat{b}_{k,\mathcal{S}}(\mathbf{x})$ has certain problems, too. For example, suppose that all realizations in $\mathcal{S}$ contain an impulse in a fixed small interval $\mathcal{K}$ but $q(b_k) < 1/2$ for all $k \in \mathcal{K}$, i.e., there is no $k \in \mathcal{K}$ at which the majority of realizations has an impulse. Then $\hat{b}_{k,\mathcal{S}}(\mathbf{x}) = 0$ for all $k \in \mathcal{K}$, which is clearly counter-intuitive since the empirical probability that there is no impulse in $\mathcal{K}$ is zero (all realizations in $\mathcal{S}$ have an impulse in $\mathcal{K}$).

The proposed block detector. We therefore use a *MAP block detector* that is a compromise between the sequence and component detectors and largely avoids their problems. The sequence $\mathbf{b}$ is split into J nonoverlapping blocks $\boldsymbol{\beta}_j$ of generally different lengths K_j, $j = 1, \dots, J$, and each block is detected independently of the others:

$$\hat{\boldsymbol{\beta}}_{j,\mathrm{MAP}}(\mathbf{x}) \triangleq \arg \max_{\boldsymbol{\beta}_j \in \{0,1\}^{K_j}} p(\boldsymbol{\beta}_j|\mathbf{x}) = \arg \max_{\boldsymbol{\beta}_j \in \mathcal{C}_j} p(\boldsymbol{\beta}_j|\mathbf{x}).$$

$$(8)$$

Here, $p(\boldsymbol{\beta}_j|\mathbf{x})$ is a marginal of $p(\mathbf{b}|\mathbf{x})$ and $\mathcal{C}_j$ is the set of all $\boldsymbol{\beta}_j \in \{0, 1\}^{K_j}$ satisfying the minimum distance condition. The sample-based version of $\hat{\boldsymbol{\beta}}_{j,\mathrm{MAP}}(\mathbf{x})$ is given by

$$\hat{\boldsymbol{\beta}}_{j,\mathcal{S}}(\mathbf{x}) \triangleq \arg \max_{\boldsymbol{\beta}_j \in \{0,1\}^{K_j}} q(\boldsymbol{\beta}_j),$$

where $q(\boldsymbol{\beta}_j)$ is the relative multiplicity of $\boldsymbol{\beta}_j$ in $\mathcal{S}$. The overall detection result is then $\hat{\mathbf{b}}_\mathcal{S}^{\mathrm{B}}(\mathbf{x}) \triangleq \left(\hat{\boldsymbol{\beta}}_{1,\mathcal{S}}^{\mathrm{T}}(\mathbf{x}) \cdots \hat{\boldsymbol{\beta}}_{J,\mathcal{S}}^{\mathrm{T}}(\mathbf{x})\right)^{\mathrm{T}}$.

To avoid problems similar to those of the component detector, the block lengths K_j should not be chosen too small. On the other hand, each K_j should be small enough so that $q(\boldsymbol{\beta}_j)$ is a good approximation to $p(\boldsymbol{\beta}_j|\mathbf{x})$; this requires that the number of hypotheses $\boldsymbol{\beta}_j \in \mathcal{C}_j$ in (8) is much smaller than M. To define the block intervals, we propose to first calculate $q(b_k = 1)$ for $k = 0, \dots, K-1$. The resulting sequence consists of "zero intervals" and "nonzero intervals."

Within zero intervals, $q(b_k = 1) \equiv 0$, and thus the sample-based block detector will always yield the zero block. Nonzero intervals, on the other hand, have $q(b_k = 1) > 0$ at all positions. They can be split into smaller intervals if they are too large, or joined with neighboring intervals if they are too small. However, we often obtained good results by directly using them as blocks.

4.2. Estimation of Amplitudes, Fringe, and Noise Variance

For estimating the complex amplitudes a_k at the positions k where an impulse was detected, we ideally use the minimum mean square error (MMSE) estimator of a_k under the condition $b_k = 1$,

$$\hat{a}_{k,\mathrm{MMSE}}(\mathbf{x}) \triangleq \mathrm{E}\{a_k | \mathbf{x}, b_k = 1\} = \int a_k \, p(a_k | \mathbf{x}, b_k = 1) \, \mathrm{d}a_k \,.$$

Calculating $p(a_k | \mathbf{x}, b_k = 1)$ by marginalization of $p(\mathbf{b}, \mathbf{a}, \boldsymbol{\alpha}, \sigma_n^2 | \mathbf{x})$ is not feasible, so we use the sample-based approximation

$$\hat{a}_{k,\mathcal{S}}(\mathbf{x}) \triangleq \frac{1}{|\mathcal{M}_k|} \sum_{m \in \mathcal{M}_k} a_k^{(m)}.$$

Here, $a_k^{(m)}$ is the amplitude of realization $(\mathbf{b}, \mathbf{a}, \boldsymbol{\alpha}, \sigma_n^2)^{(m)}$ at position k (i.e., the kth entry of the $\mathbf{a}$ contained in $(\mathbf{b}, \mathbf{a}, \boldsymbol{\alpha}, \sigma_n^2)^{(m)}$) and $\mathcal{M}_k$ is the set of indices m of all realizations with $b_k^{(m)} = 1$. We calculate $\hat{a}_{k,\mathcal{S}}(\mathbf{x})$ only at those k where an impulse was detected, i.e., where $\left[\hat{\mathbf{b}}_{\mathcal{S}}^{\mathrm{B}}(\mathbf{x})\right]_k = 1$ (this also guarantees that $\mathcal{M}_k$ is nonempty).

Finally, sample-based versions of the MMSE estimators of the fringe coefficient vector $\boldsymbol{\alpha}$ and the noise variance σ_n^2 are given by

$$\hat{\boldsymbol{\alpha}}_{\mathcal{S}}(\mathbf{x}) \triangleq \frac{1}{M} \sum_{m=1}^{M} \boldsymbol{\alpha}^{(m)}, \qquad \widehat{\sigma_{n,\mathcal{S}}^2}(\mathbf{x}) \triangleq \frac{1}{M} \sum_{m=1}^{M} (\sigma_n^2)^{(m)},$$

where $\boldsymbol{\alpha}^{(m)}$ and $(\sigma_n^2)^{(m)}$ are, respectively, the $\boldsymbol{\alpha}$ and σ_n^2 of realization $(\mathbf{b}, \mathbf{a}, \boldsymbol{\alpha}, \sigma_n^2)^{(m)}$.

5. GIBBS SAMPLER

For drawing a sample $\mathcal{S} = \{(\mathbf{b}, \mathbf{a}, \boldsymbol{\alpha}, \sigma_n^2)^{(m)}\}_{m=1,\dots,M}$ from the joint posterior $p(\mathbf{b}, \mathbf{a}, \boldsymbol{\alpha}, \sigma_n^2 | \mathbf{x})$, we use the Gibbs sampler [10]. Assuming a generic multivariate distribution $p(\mathbf{z})$, the Gibbs sampler draws a sample $\{\mathbf{z}^{(m)}\}_{m=1,\dots,M}$ from $p(\mathbf{z})$ in a recursive fashion. A new realization $\mathbf{z}^{(m+1)}$ is obtained from the current realization $\mathbf{z}^{(m)}$ as follows. Each entry $z_k^{(m+1)}$ is drawn from the univariate distribution $p(z_k | \mathbf{z}_{\sim k})$, where $\mathbf{z}_{\sim k}$ denotes $\mathbf{z}$ without entry z_k and the values used in the condition $\mathbf{z}_{\sim k}$ are the most recent ones (i.e., in general, some of them have already been updated). The stationary distribution of the resulting Markov chain $\mathbf{z}^{(m)}$, $m = 0, 1, \dots$ converges to $p(\mathbf{z})$ as $M \to \infty$ [10]. Thus, the sample $\{\mathbf{z}^{(m)}\}_{m=1,\dots,M}$ will represent $p(\mathbf{z})$ if it is large enough and if an initial "burn-in period" (which is affected by the initialization $\mathbf{z}^{(0)}$) is removed.

The univariate distributions $p(z_k | \mathbf{z}_{\sim k})$ can be derived from the likelihood function (in our case (5)) and the priors (see Section 3). Because of space limitations, we will present them without proof.

Amplitudes. Using the fact that $p(a_k)$ is a conjugate prior [11], we obtain $p(a_k | \mathbf{x}, \mathbf{b}, \mathbf{a}_{\sim k}, \boldsymbol{\alpha}, \sigma_n^2) = \mathcal{CN}(\mu, \sigma^2)$, with

$$\mu = \frac{\sigma^2 b_k \mathbf{f}_k^{\mathrm{H}}(\mathbf{x} - \mathbf{F}\mathbf{B}_{\sim k}\mathbf{a}_{\sim k})}{\sigma_n^2}, \quad \sigma^2 = \left(\frac{b_k^2 \|\mathbf{f}_k\|^2}{\sigma_n^2} + \frac{1}{\sigma_a^2}\right)^{-1}. \tag{9}$$

Here, $\mathbf{f}_k$ denotes the kth column of $\mathbf{F}$ and $\mathbf{B}_{\sim k}$ denotes $\mathbf{B}$ without the kth column.

Impulse indicator. For b_k, we find

$$p(b_k | \mathbf{x}, \mathbf{b}_{\sim k}, \mathbf{a}_{\sim k}, \boldsymbol{\alpha}, \sigma_n^2) = C \sigma^2 e^{|\mu|^2/\sigma^2} p(\mathbf{b}), \tag{10}$$

with μ and σ^2 as defined in (9). The constant C is easily calculated since $b_k \in \{0, 1\}$. However, using (10), many Gibbs sampler iterations are needed to obtain changes in $\mathbf{b}$. This is because due to the minimum distance condition, b_k is almost determined by its neighborhood. (A similar phenomenon was observed in [12, 13].) We will thus use a modified distribution whose condition ignores not just b_k but all $b_{k'}$ within a neighborhood of length $d_{\min}$ of position k, on the side at which we sample. This neighborhood is defined by $\mathcal{K} \triangleq [k, k_{\max}]$ with $k_{\max} \triangleq \min\{k + d_{\min} - 1, K\}$ (we assume that the $b_{k'}$ for $k' < k$ have already been sampled). We then use the Gibbs sampler with $p(b_k | \mathbf{x}, \mathbf{b}_{\sim k}, \mathbf{a}_{\sim k}, \boldsymbol{\alpha}, \sigma_n^2)$ replaced by $p(b_k | \mathbf{x}, \mathbf{b}_{\sim \mathcal{K}}, \mathbf{a}_{\sim \mathcal{K}}, \boldsymbol{\alpha}, \sigma_n^2)$, where, e.g., $\mathbf{b}_{\sim \mathcal{K}}$ denotes $\mathbf{b}$ without the entries indexed by $\mathcal{K}$. This latter pmf can be calculated as $p(b_k | \mathbf{x}, \mathbf{b}_{\sim \mathcal{K}}, \mathbf{a}_{\sim \mathcal{K}}, \boldsymbol{\alpha}, \sigma_n^2) = \sum_{\boldsymbol{\beta}_{\sim k}} p(\boldsymbol{\beta} | \mathbf{x}, \mathbf{b}_{\sim \mathcal{K}}, \mathbf{a}_{\sim \mathcal{K}}, \boldsymbol{\alpha}, \sigma_n^2)$, where $\boldsymbol{\beta} \triangleq (b_k \ b_{k+1} \cdots b_{k_{\max}})^{\mathrm{T}}$ comprises all $b_{k'}$ indexed by $\mathcal{K}$. Note that there are only $d_{\min} + 1$ different $\boldsymbol{\beta}$ because in $\mathcal{K}$ there can be at most one impulse. One can show that $p(\boldsymbol{\beta} | \mathbf{x}, \mathbf{b}_{\sim \mathcal{K}}, \mathbf{a}_{\sim \mathcal{K}}, \boldsymbol{\alpha}, \sigma_n^2)$ is still given by expression (10), however with

$$\mu = \frac{\sigma^2 \boldsymbol{\beta}^{\mathrm{H}} \mathbf{F}_{\mathcal{K}}^{\mathrm{H}}(\mathbf{x} - \mathbf{F}_{\sim \mathcal{K}} \mathbf{B}^{\sim \mathcal{K}} \mathbf{a}_{\sim \mathcal{K}})}{\sigma_n^2}$$

$$\sigma^2 = \left(\frac{\boldsymbol{\beta}^{\mathrm{H}} \mathbf{F}_{\mathcal{K}}^{\mathrm{H}} \mathbf{F}_{\mathcal{K}} \boldsymbol{\beta}}{\sigma_n^2} + \frac{1}{\sigma_a^2}\right)^{-1}.$$

Here, $\mathbf{F}_{\mathcal{K}}$ consists of the columns of $\mathbf{F}$ indexed by $\mathcal{K}$, $\mathbf{F}_{\sim \mathcal{K}}$ denotes $\mathbf{F}$ without these columns, and $\mathbf{B}^{\sim \mathcal{K}} \triangleq \mathrm{diag}(\mathbf{b}_{\sim \mathcal{K}})$.

Fringe. Because $\boldsymbol{\alpha}$ is Gaussian, it can be treated as one parameter in the Gibbs sampler without incurring excessive complexity. The conjugate prior property yields $p(\boldsymbol{\alpha} | \mathbf{x}, \mathbf{b}, \mathbf{a}, \sigma_n^2) = \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where (recall $\mathbf{R} = \mathrm{toep}(\mathbf{r})$ and (3))

$$\boldsymbol{\mu} = \frac{\boldsymbol{\Sigma} \mathbf{H}^{\mathrm{H}} \mathbf{R}^{\mathrm{H}} \mathbf{x}}{\sigma_n^2}, \quad \boldsymbol{\Sigma} = \left(\frac{\mathbf{H}^{\mathrm{H}} \mathbf{R}^{\mathrm{H}} \mathbf{R} \mathbf{H}}{\sigma_n^2} + \frac{\mathbf{I}}{\sigma_\alpha^2}\right)^{-1}.$$

The time-shift ambiguity inherent to the convolution model is resolved as discussed in [2].

Noise variance. Since $p(\sigma_n^2)$ in (4) is a conjugate prior, we obtain $p(\sigma_n^2 | \mathbf{x}, \mathbf{b}, \mathbf{a}, \boldsymbol{\alpha}) = \mathcal{IG}(\xi + K, \eta + \|\mathbf{x} - \mathbf{F}\mathbf{B}\mathbf{a}\|^2)$.

6. NUMERICAL RESULTS

We next demonstrate the performance of our method and compare it with that of the Gibbs sampler based blind deconvolution method of [2], hereafter referred to as "reference method" (RM).

Synthetic signals. We simulated 100 RFs $r_k = a_k b_k$ and corresponding noisy time-domain signals $x_k = r_k * f_k + n_k$, using a fixed fringe with Gaussian shape and height 1 (see the right-hand plots in Fig. 1(b), (c)), a fixed noise variance σ_n^2, and pmf/pdf parameters $p_0 = 0.96$, $d_{\min} = 12$, and $\sigma_a^2 = 7.74 \cdot 10^{-5}$. We chose σ_n^2 to obtain an SNR ($\|\mathbf{Fr}\|^2/\|\mathbf{n}\|^2$) of 16 dB. Both methods used the true p_0, $d_{\min}$, σ_a^2 and, in (4), $\eta = 0.5$ and the value of ξ for which the mean $\eta/(\xi-1)$ equals the true σ_n^2. We used $L = 7$ Hermite functions in our method and a fringe length of 33 in the RM.

Fig. 1 shows a typical example. While the results of both methods are seen to be satisfactory (with some amplitude estimation errors and spurious impulses in the RM result), our method requires only three iterations to achieve a better result than does the RM with 4000 iterations (the number of iterations used in [2]). Since one iteration of our method is similarly complex as four iterations of the RM, our three iterations correspond to about 12 RM iterations. However, the RM results after 12 iterations (not shown) are still extremely poor.
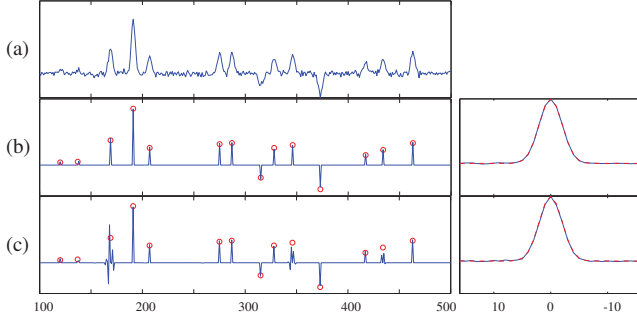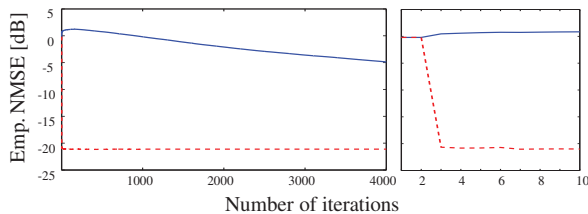
Fig. 1. Results for a synthetic signal: (a) Signal, (b) RF estimate (left) and fringe estimate (right) obtained with the proposed method after three iterations, (c) RF estimate (left) and fringe estimate (right) obtained with the reference method after 4000 iterations. The circles indicate the true impulse positions and amplitudes. The dashed lines in the right-hand plots (almost coinciding with the solid lines) indicate the true fringe function. Real parts are shown throughout.

Fig. 2 shows the empirical normalized mean-square error (NMSE) of RF estimation versus the number of iterations. The empirical NMSE is the average (over 100 realizations) of $\|\hat{\mathbf{r}} - \mathbf{r}\|^2$ normalized by the average of $\|\mathbf{r}\|^2$. In our method, we resynthesized $\hat{\mathbf{r}}$ from the detected impulses and estimated amplitudes. (Note that the RM directly yields $\hat{\mathbf{r}}$.) The number of iterations equals the total length of the Markov chain. Out of each chain, the last 25% of the iterations were used to constitute the sample. It is seen that the RM converges very slowly, whereas our method achieves an NMSE below -20 dB after only three iterations. We note that an improvement of the RM resulting in faster convergence was proposed very recently [13].

Real OCT signal. We applied the two methods to a real OCT signal measured by a frequency-domain OCT device (frequency range 320.7–425.7 THz) scanning a human retina. The parameters (chosen for best performance) were $p_0 = 0.85$, $d_{\min} = 7$, $\sigma_a^2 = 3.91 \cdot 10^{-5}$, SNR $= 26$ dB, and $L = 30$ for our method and $p_0 = 0.973$, $\sigma_a^2 = 8.79 \cdot 10^{-5}$, SNR $= 16$ dB, and fringe length 33 for the RM. Fig. 3(a) shows a segment of the time-domain signal corresponding to a length of $530\,\mu$m. From Fig. 3(b),(c), it is seen that the RF estimate obtained with our method after 20 iterations is sparse whereas that obtained with the RM after 4000 iterations is not. The two fringe estimates are quite different; a validation of these estimates based on fringe measurements will be part of our future work.

7. CONCLUSION

We proposed a Gibbs sampler method for inferring the depths and reflection coefficients of tissue interfaces from signals measured by frequency-domain OCT instrumentation. Knowledge of the instrumentation-dependent fringe function is not required. Our method enforces sparsity of the tissue interfaces by using an impulse detector

Fig. 2. Empirical NMSE of the RF estimate (dashed line: proposed method, solid line: reference method) versus the number of iterations. Left: all 4000 iterations, right: the first 10 iterations.
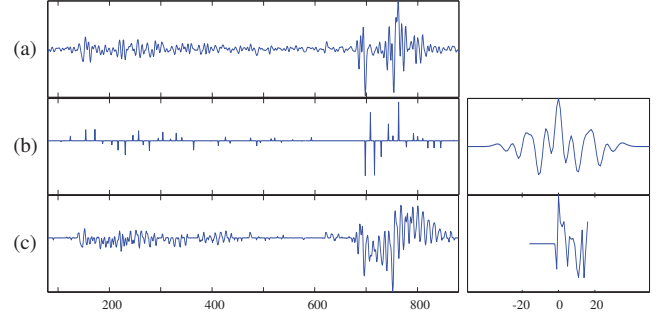
Fig. 3. Results for a real OCT signal: (a) Signal, (b) RF estimate (left) and fringe estimate (right) obtained with the proposed method after 20 iterations, (c) RF estimate (left) and fringe estimate (right) obtained with the reference method after 4000 iterations. Real parts are shown throughout.

and a modified Bernoulli-Gaussian prior with a minimum distance constraint. Numerical results demonstrated significant performance gains compared to the blind deconvolution method proposed in [2].

8. ACKNOWLEDGMENT

9. REFERENCES

[1] A. F. Fercher, W. Drexler, C. K. Hitzenberger, and T. Lasser, "Optical Coherence Tomography—Principles and applications," *Rep. Progr. Phys.*, vol. 66, pp. 239–303, Feb. 2003.

[2] Q. Cheng, R. Chen, and T.-H. Li, "Simultaneous wavelet estimation and deconvolution of reflection seismic signals," *IEEE Trans. Geosc. Remote Sens.*, vol. 34, pp. 377–384, Mar. 1996.

[3] J. J. Kormylo and J. M. Mendel, "Maximum likelihood detection and estimation of Bernoulli-Gaussian processes," *IEEE Trans. Inf. Theory*, vol. IT-28, pp. 482–488, May 1982.

[4] C. Novak, "Optical Coherence Tomography: Signal modeling and processing," Master's thesis, Vienna University of Technology, Vienna, Austria, 2006.

[5] S. C. Sekhar, R. A. Leitgeb, M. L. Villiger, A. H. Bachmann, T. Blu, and M. Unser, "Non-iterative exact signal recovery in frequency domain optical coherence tomography," in *Proc. IEEE ISBI-07*, (Metro Washington, DC), pp. 808–811, April 2007.

[6] A. J. E. M. Janssen, "Positivity and spread of bilinear time-frequency distributions," in *The Wigner Distribution — Theory and Applications in Signal Processing* (W. Mecklenbräuker and F. Hlawatsch, eds.), pp. 1–58, Amsterdam, The Netherlands: Elsevier, 1997.

[7] F. Hlawatsch, *Time-Frequency Analysis and Synthesis of Linear Signal Spaces: Time-Frequency Filters, Signal Detection and Estimation, and Range-Doppler Estimation*. Boston (MA): Kluwer, 1998.

[8] N. Dobigeon, J.-Y. Tourneret, and M. Davy, "Joint segmentation of piecewise constant autoregressive processes by using a hierarchical model and a Bayesian sampling approach," *IEEE Trans. Signal Processing*, vol. 55, pp. 1251–1263, Apr. 2007.

[9] O. Rabaste and T. Chonavel, "Estimation of multipath channels with long impulse response at low SNR via an MCMC method," *IEEE Trans. Signal Processing*, vol. 55, pp. 1312–1325, Apr. 2007.

[10] J. C. Spall, "Estimation via Markov chain Monte Carlo," *IEEE Contr. Syst. Mag.*, pp. 34–45, Apr. 2003.

[11] C. P. Robert, *The Bayesian Choice*. New York: Springer, 1996.

[12] S. Bourguignon and H. Carfantan, "Bernoulli-Gaussian spectral analysis of unevenly spaced astrophysical data," in *IEEE Workshop on Statistical Signal Processing*, (Bordeaux, France), pp. 811–816, July 2005.

[13] D. Ge, J. Idier, and E. Le Carpentier, "A new algorithm for blind Bernoulli-Gaussian deconvolution," in *Proc. EUSIPCO-08*, (Lausanne, Switzerland), Aug. 2008.